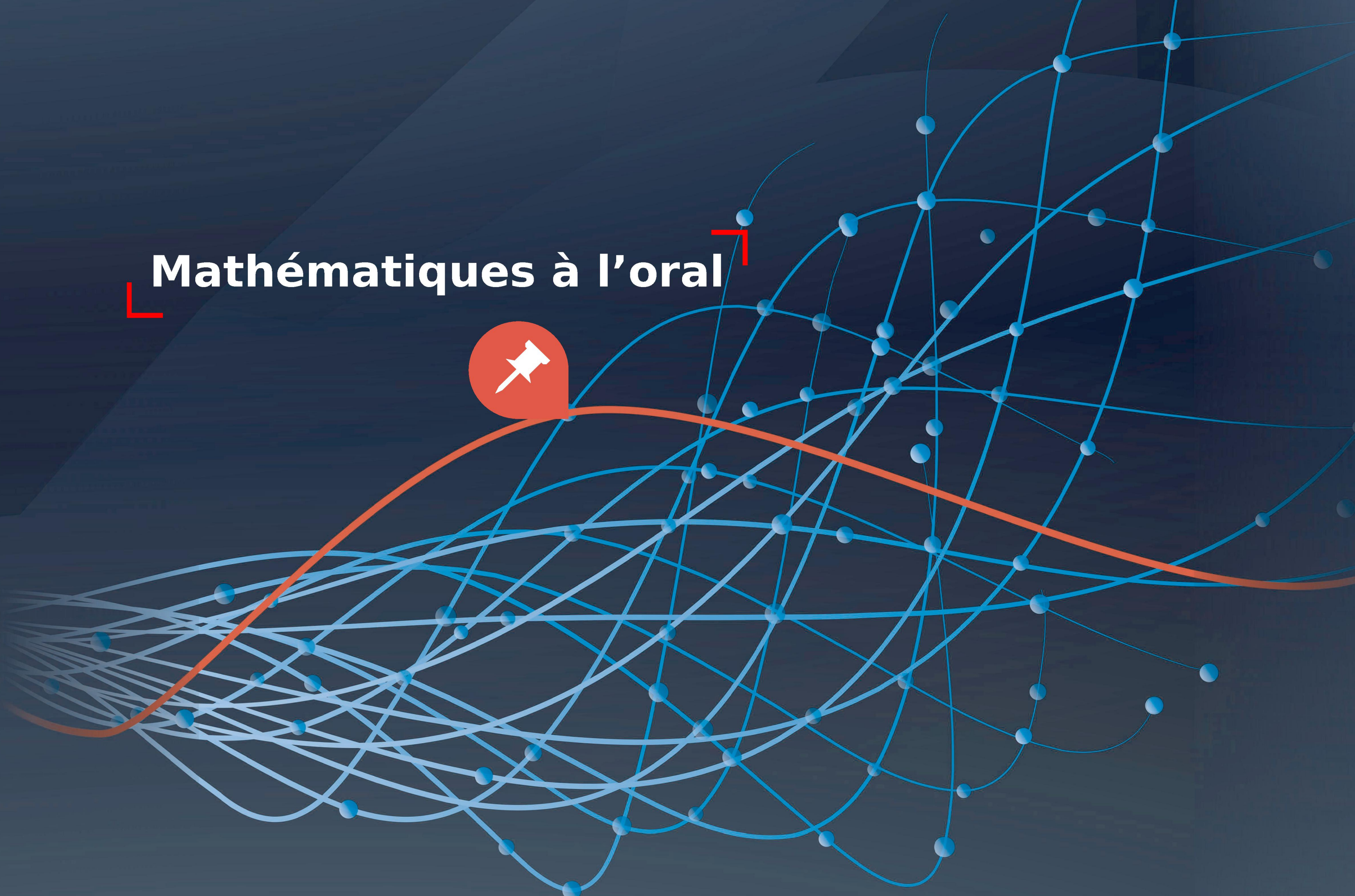
Le bulletin de l'APMEP - N° 538

AU FIL DES MATHS

de la maternelle à l'université. . .

Édition Octobre, Novembre, Décembre 2020

Mathématiques à l'oral



APMEP

Association des Professeurs de Mathématiques de l'Enseignement Public

ASSOCIATION
DES PROFESSEURS DE MATHÉMATIQUES
DE L'ENSEIGNEMENT PUBLIC

26 rue Duméril, 75013 Paris
Tél. : 01 43 31 34 05 - Fax : 01 42 17 08 77
Courriel : secretariat-apmep@orange.fr - Site : https://www.apmep.fr
Présidente d'honneur : Christiane ZEHREN

Au fil des maths, c'est aussi une revue numérique augmentée :
https://afdm.apmep.fr
version réservée aux adhérents. Pour y accéder connectez-vous à votre compte *via* l'onglet *Au fil des maths* (page d'accueil du site) ou *via* le QRcode, ou suivez les logos.

Si vous désirez rejoindre l'équipe d'*Au fil des maths* ou bien proposer un article, écrivez à aufildesmaths@apmep.fr

À ce numéro est joint le BGV nº 215

ÉQUIPE DE RÉDACTION

Votre adhésion à l'APMEP vous abonne automatiquement à *Au fil des maths*.
Pour les établissements, le prix de l'abonnement est de 60 € par an.
La revue peut être achetée au numéro au prix de 15 € sur la boutique en ligne de l'APMEP.

Mise en page : François PÉTIARD
Dépôt légal : Décembre 2020
Impression : Imprimerie Corlet.
ZI, rue Maximilien Vox BP 86, 14110 Condé-sur-Noireau ISSN : 2608-9297

APMEP

Mathématiques et systèmes de retraite

Toute réflexion sur les systèmes de retraites incite à nous questionner sur les notions d'égalité et de justice dans la répartition des retraites. Pierre Carriquiry nous présente quelques outils et modèles mathématiques pour le faire.

Pierre Carriquiry

Introduction

Toute réforme des systèmes de retraites fait émerger des mathématiques intéressantes pour répondre à deux grandes questions : comment prélever la somme qui sera distribuée aux retraités ? Comment distribuer cette somme ? Bien entendu, les mathématiques ne suffisent pas à répondre à ces questions et inévitablement les questions politiques de justice, d'égalité et d'équité entrent en jeu mais ce faisant, elles offrent aux mathématiques un nouveau champ d'application : comment définir et mesurer des inégalités dans les différents modèles choisis ? Dans cet article, nous donnons des éléments de réflexion uniquement sur la question de la distribution de la somme et laissons de côté la question du prélèvement.

Le principe de répartition

Dans un système de retraite par répartition, les retraites d'une année sont financées par les cotisations des actifs de la même année ou par les impôts. Dans la suite on supposera que la somme disponible pour payer les retraites d'une année est fixée ; on la notera s. Elle peut être égale à la somme des cotisations des actifs de la même année ou à un pourcentage du PIB par exemple. Notre question centrale est donc : « comment répartir la somme s entre les n retraités ? » Mathématiquement, il s'agit de résoudre une équation du type $x_1 + \cdots + x_n = s$ lorsque les inconnues (les montants x_i des retraites) sont soumises à un certain nombre de contraintes. On va définir quelques méthodes de répartition, les questions d'égalité que ces méthodes posent et étudier les problèmes mathématiques de mesure des inégalités des systèmes de retraite qui en découlent.

La répartition égalitaire

Principe

Il s'agit de l'égalité au sens purement mathématique du terme : toutes les retraites sont égales. Il y a une seule solution : chaque retraité aura une retraite égale à $m = \dfrac{s}{n}$.

Exemple

Dans un pays de 15 millions de retraités, si $s = 300$ milliards d'euros, chaque retraité touche 20 000 euros par an. Toute autre répartition est inégalitaire.

Égalité par groupe

Si un groupe de retraités exige de percevoir plus de 20 000 euros par membre de son groupe, il faudra trouver d'autres groupes qui accepteront de recevoir moins de 20 000 euros. Par exemple, si un groupe de 5 millions de retraités exige de toucher 30 000 euros, les 10 millions qui ne font pas partie de ce groupe auront une retraite de $\dfrac{300 \times 10^9 - 5 \times 10^6 \times 30\,000}{10^7} = 15\,000$ euros en supposant une répartition égalitaire au sein de chaque groupe.

Plus généralement si l'ensemble des retraités E se compose de k sous-ensembles disjoints $E_1, \ldots, E_k$, d'effectifs $n_1, \ldots, n_k$ on sait que la moyenne m des retraites dans la population E sera la moyenne (pondérée par les effectifs) des moyennes $m_1, \ldots, m_k$ dans les sous-ensembles. On a donc la relation $m = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i m_i$.

Ainsi comme $m = \dfrac{s}{n}$ est fixé, si chaque groupe exige une méthode de calcul spécifique pour sa retraite, ce qui revient à fixer les moyennes $m_1, \ldots, m_k$, il est peu vraisemblable que l'égalité $m = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i m_i$ soit vérifiée.

On aboutit ainsi à une première question politique : dans un pays où la notion d'égalité figure dans plusieurs articles de la Constitution, tout groupe de retraités qui exige de toucher une retraite supérieure à la moyenne devrait, soit trouver d'autres groupes qui acceptent de percevoir une retraite inférieure à la moyenne, soit justifier sa revendication.

Systèmes par points

La plupart des justifications d'un système peuvent se traduire sous la forme suivante : « *pendant ma vie active j'ai eu une utilité sociale qui peut être quantifiée par l'attribution de points et ma retraite doit être proportionnelle au total des points obtenus lorsque je la prends.* »

Principe

Chaque personne qui prend sa retraite obtient un certain nombre de points. Soit T le nombre total de points acquis par tous les retraités d'une année ; si p est le nombre de points acquis par un retraité, sa retraite sera $\frac{sp}{T}$.
Comment attribuer ces points ?
Les deux principales justifications de l'inégalité (au sens mathématique) d'un système de retraite sont :

1. la quantité ou la pénibilité du travail effectué pendant la vie active ;
2. le montant des cotisations payées pendant la vie active.

On va proposer quelques systèmes par points qui tiennent plus ou moins compte de ces critères.

Quelques systèmes

- *Retraites en fonction du travail effectué :*
  Une heure de travail donne un point ou $(1+\epsilon)$ point pour tenir compte de la pénibilité. Les femmes (ou les hommes) qui ont élevé des enfants sans percevoir une rémunération auront des points correspondant au temps consacré à cette tâche.
- *Retraites en fonction des cotisations versées :*
  Chaque euro cotisé donne un point. Ce système suppose que chaque travailleur verse des cotisations retraite sur les revenus qu'il perçoit.

- *Retraites en fonction des impôts payés :*
  Chaque euro d'impôt payé durant la vie active donne un point. Ce système suppose qu'il n'y a pas de cotisations retraite, que les retraites sont financées par les impôts et que tous les travailleurs paient des impôts.
- *Retraites en fonction des meilleurs revenus perçus :*
  Le nombre de points d'un travailleur qui prend sa retraite est égal au meilleur revenu annuel de sa vie active.

Choix d'un système

Il est clair que le premier système est plus égalitaire que les trois autres systèmes qui reproduisent plus ou moins les inégalités des revenus des actifs et qui semblent classés par ordre d'inégalité croissante. Est-il normal de reproduire ces inégalités à la retraite ? Les partisans de ces systèmes diront que ceux qui ont plus cotisé, ou payé plus d'impôts, ou gagné plus ont droit à une retraite plus élevée. On peut leur répondre que les différences de salaires ne sont pas toujours justifiées et si on les justifie par la loi de l'offre et de la demande (les salaires sont plus élevés dans les professions où on trouve peu de candidats) cette loi n'a aucune raison de s'appliquer aux retraités car il n'y a pas de demande pour les retraités.

Si on justifie l'inégalité des salaires par l'utilité sociale des professions (encore faut-il s'entendre sur la notion d'utilité sociale), la seule justification de l'inégalité des retraites est alors que les retraités qui ont eu une utilité sociale plus grande doivent être récompensés par une retraite plus élevée. Mais les périodes d'épidémie nous montrent que cette utilité sociale peut être très variable : une aide-soignante par exemple a alors une utilité sociale bien plus grande que celle de bien d'autres professions. De plus, l'utilité sociale des retraités peut être beaucoup plus complexe à évaluer par la multiplicité et la variabilité de leurs activités. Par ailleurs, tous ces systèmes par points ont un inconvénient : les retraités qui n'ont pas travaillé (ou peu travaillé) ne touchent rien (ou presque rien). On peut alors décider d'accorder une retraite minimale r_m à tous. La formule de calcul doit être modifiée. Exemple : on veut répartir une somme de 120 (milliers d'euros) entre cinq retraités ayant obtenu $(15, 16, 18, 19, 32)$ points. Si on répartit la somme totale proportionnellement aux nombres de points, les retraites seront respectivement (en milliers d'euros) : 18 ; 19,2 ; 21,6 ; 22,8 ; 38,4. Si on décide que la retraite minimale est 22 milliers d'euros, il serait faux de penser que seuls les trois premiers toucheront cette retraite minimale. En effet si on accorde une retraite de 22 aux trois premiers, il reste $120 - 66 = 54$ pour les deux autres et, si on partage cette somme proportionnellement aux nombres de points $(19, 32)$, ils toucheront respectivement $54 \times \dfrac{19}{51} = 20{,}12$ (soit moins que la retraite minimale 22) et $54 \times \dfrac{32}{51} = 33{,}88$. Il faudra alors accorder la retraite minimale à quatre retraités et le dernier prendra ce qui reste. Dans le cas général, soit $(p_1, \ldots, p_n)$, les nombres de points des n retraités rangés dans l'ordre croissant, k_0 le nombre de retraités qui n'ont pas en principe assez de points pour toucher la retraite minimale r_m ; le nombre k de retraités qui toucheront la retraite minimale doit vérifier : $k_0 \leqslant k$ et $\forall\, i \in \{k+1, \ldots, n\}$, $(s - kr_m)\dfrac{p_i}{\sum_{i=k+1}^{n} p_i} > r_m$.

L'instauration d'une retraite minimale réduit l'inégalité du système mais comment mesurer cette réduction et même comment définir cette inégalité ? Les mathématiques nous offrent plusieurs outils pour ce faire.

Mesures de l'inégalité

Déciles

Soit D_1, ..., D_9 les déciles de la distribution des retraites ($10k$ % des retraités gagnent moins que D_k). Pour mesurer l'inégalité de la répartition, on peut utiliser les rapports $\frac{D_9}{D_1}$, $\frac{D_9}{D_5}$, $\frac{D_5}{D_1}$ etc. Dans la répartition égalitaire, tous les déciles sont égaux et ces rapports sont égaux à 1. Dans les autres répartitions ils sont supérieurs ou égaux à 1 et plus ils sont grands plus l'inégalité est forte. Par exemple $\frac{D_9}{D_1} = 20$ signifie : 10 % des retraités gagnent moins de D_1, 10 % gagnent plus de D_9 et la retraite D_9 est 20 fois plus grande que la retraite D_1. L'inconvénient de ces indices est qu'ils ne tiennent pas compte de toutes les valeurs.

Variance (ou écart-type)

Recherche d'indices utilisant la variance

Soit $x_1, \ldots, x_n$ les retraites des n retraités rangées dans l'ordre croissant, $m = \frac{s}{n}$ la moyenne arithmétique des retraites. La variance $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$ est d'autant plus grande que les retraites s'écartent de la moyenne et peut donc être utilisée pour mesurer l'inégalité du système. L'inconvénient est qu'elle s'exprime en euros², mais on peut utiliser l'écart-type qui s'exprime en euros. Cependant, on préfère avoir un indice sans dimension de préférence compris entre 0 et 1 qui peut être interprété comme un pourcentage. On peut alors utiliser $\frac{V}{V_{\max}}$ où $V_{\max}$ est le maximum de la variance de n nombres choisis dans l'intervalle $[0\,;a]$ de moyenne fixée m, a étant un nombre positif supérieur à m. On a[1] $V_{\max} = \frac{1}{n}\left[(nm - ka)^2 + ka^2\right] - m^2$ où k est la partie entière de $\frac{nm}{a}$. Par exemple pour $n = 15 \times 10^6$, $m = 20\,000$ et $a = 10^6$ (qui serait le montant maximal d'une retraite) on trouve $V_{\max} = 196 \times 10^8$ soit un écart-type maximal de 140 000 euros. Cette valeur montre l'inconvénient de l'indice $\frac{V}{V_{\max}}$, les valeurs réelles de la variance étant très inférieures à cette valeur maximale. Dans le cas d'une retraite minimale b et d'une retraite maximale a, on est amené à chercher le maximum de la variance de n nombres appartenant à l'intervalle $[b\,;a]$ de moyenne fixée m. En posant $x'_i = x_i - b$ on est ramené au cas précédent car la variance ne change pas si on ajoute une constante à chaque valeur et on obtient : $V_{\max} = \frac{1}{n}\left[(n(m-b) - k(a-b))^2 + k(a-b)^2\right] - (m-b)^2$ où k est la partie entière de $\frac{n(m-b)}{a-b}$.
L'inconvénient de cette valeur maximale est qu'elle dépend des paramètres a et b. Alors on prend $b = x_1$ (plus petite valeur) et $a = x_n$ (plus grande valeur) et on a :

$$\boxed{V_{\max}=\frac{1}{n}\left[(n(m-x_1) - k(x_n-x_1))^2 + k(x_n-x_1)^2\right] - (m-x_1)^2 \text{ où } k \text{ est la partie entière de } \frac{n(m-x_1)}{x_n-x_1}.}$$

L'indice $I = \frac{V}{V_{\max}}$ est alors un indice qui mesure l'inégalité de la répartition et qui peut être interprété comme un pourcentage. De plus cet indice possède deux propriétés intéressantes : si on ajoute une constante à toutes les valeurs on sait que la variance V ne change pas et on voit que $V_{\max}$ est aussi invariant car les différences $(x_n - x_1)$ et $(m - x_1)$ ne changent pas, donc $I = \frac{V}{V_{\max}}$ est invariant. Si on multiplie toutes les valeurs par une constante c, on sait que la variance V est multipliée par c^2 et on voit que $V_{\max}$ est aussi multiplié par c^2, donc $I = \frac{V}{V_{\max}}$ est invariant par changement d'unité (ici monétaire).

1. Voir la démonstration dans l'article de S. Chesney dans le Bulletin Vert nº 443 (décembre 2002) page 740 ou dans le courrier des lecteurs du Bulletin Vert nº 447 (octobre 2003) page 524.

Exemple : pour $n = 15 \times 10^6$; $x_1 = 12\,000$; $x_n = 132\,000$; $m = 20\,000$; on trouve $V_{\max} = 896 \times 10^6$ et si la variance observée est $15\,000^2$ on a : $I = 0{,}25$ c'est-à-dire que la variance observée est égale à 25 % de la variance maximale et l'écart-type observé est égal à 50 % de l'écart-type maximal qui correspond à la répartition suivante : un million de retraités touchent la retraite maximale 132 000 et 14 millions touchent la retraite minimale 12 000 ce qui serait évidemment très inégalitaire.

Dans la répartition égalitaire, $V = 0$ et réciproquement. On peut alors convenir que $I = \dfrac{V}{V_{\max}} = 0$ bien que si on applique la dernière formule de $V_{\max}$ le dénominateur est aussi nul (il faudrait alors appliquer la formule dans laquelle on définit arbitrairement une retraite minimale b et une retraite maximale a).

Si I est proche de zéro, les valeurs sont groupées autour de la moyenne, et si I est proche de 1 on observe beaucoup de valeurs proches de la retraite maximale ou de la retraite minimale.

Indices mesurant l'inégalité entre des groupes

Dans la réalité, les méthodes de calcul des retraites sont différentes suivant des groupes d'individus ayant en général la même profession. On est alors amené à mesurer les inégalités entre les groupes. Dans le cas où la population E des retraités est divisée en k sous-ensembles $E_1, \ldots, E_k$ disjoints, on peut calculer les moyennes $m_1, \ldots, m_k$ et les variances $v_1, \ldots, v_k$ dans chaque sous-ensemble. La variance dans la population est alors : $v = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i(m_i - m)^2 + \dfrac{1}{n}\sum_{i=1}^{k} n_i v_i$. Le terme $\dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i(m_i - m)^2$ est la variance des moyennes qui mesure l'inégalité entre les différents groupes. Le terme $\dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i v_i$ est la moyenne des variances qui mesure les inégalités à l'intérieur de chaque groupe. Exemple : dans une population de 15 millions de retraités, 5 millions ont une retraite moyenne de 30 000 euros avec une variance de $(5\,000)^2$ et 10 millions ont une retraite moyenne de 15 000 euros avec une variance de $(1\,000)^2$.

La moyenne des variances est : $\dfrac{1}{15}\left[5(5\,000)^2 + 10(1\,000)^2\right] = 9 \times 10^6$.

La variance des moyennes est : $\dfrac{1}{15}\left[5(10\,000)^2 + 10(5\,000)^2\right] = 50 \times 10^6$.

La variance des retraites est donc 59×10^6 et l'on peut dire que $\dfrac{50}{59} \approx 85$ % de cette variance s'explique par l'inégalité entre les deux groupes et que 15 % s'explique par l'inégalité à l'intérieur de chaque groupe. On voit ainsi apparaître deux moyens différents d'agir pour réduire les inégalités : soit en modifiant la répartition à l'intérieur de chacun des groupes, soit en modifiant la répartition entre les deux groupes.

Indice de Gini

La variance mesure les écarts par rapport à la moyenne. Il existe un indice qui mesure les écarts entre les différentes valeurs appelé indice de Gini. Dans la répartition égalitaire, x % des retraités perçoivent x % du montant total des retraites pour toutes les valeurs possibles de x. Dans une répartition inégalitaire, x % des retraités les plus défavorisés perçoivent moins de x % du total des retraites. L'indice de Gini mesure ces écarts.

Définition géométrique

Soit n nombres $x_1, \ldots, x_n$ rangés dans l'ordre croissant représentant les retraites, s leur somme et $m = \frac{s}{n}$ leur moyenne arithmétique.

Dans un repère orthonormé on place les points $M_0(0\ ;\ 0)$, ..., $M_k\left(\frac{k}{n}\ ;\ y_k\right)$, ..., $M_n\left(\frac{n}{n}\ ;\ y_n\right)$ où $y_k = \frac{1}{s}\sum_{i=1}^{k} x_i$ est la proportion du montant total des retraites perçues par les k retraités qui perçoivent les retraites les plus basses. Dans le système égalitaire, toutes les retraites sont égales à m, $y_k = \frac{k}{n}$ et tous les points M_k sont sur le segment $[(0;0)\,;(1;1)]$. Dans tous les autres cas, les points sont en-dessous (au sens large) de ce segment. La ligne polygonale (M_0, ..., M_n) qui joint ces points s'appelle courbe de Lorenz et l'aire du domaine limité par cette courbe et le segment $[(0;0)\,;(1;1)]$ est une mesure de l'inégalité de la répartition. Cette aire est inférieure à 0,5, alors on prend le double de cette aire appelé indice de Gini pour mesurer l'inégalité.

Formules

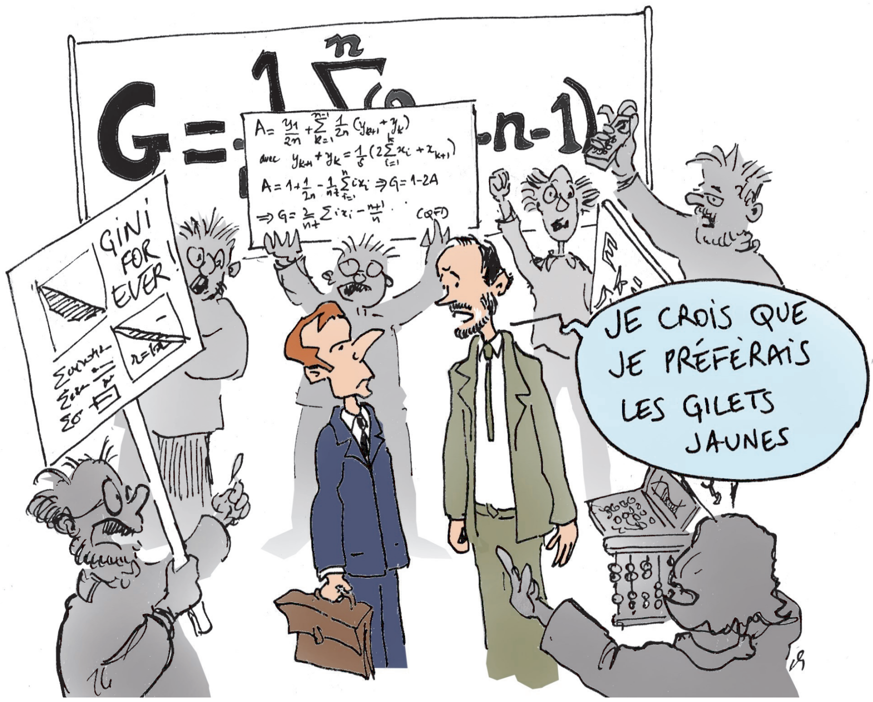
Soit A l'aire du domaine limité par l'axe des abscisses et la courbe de Lorenz. L'indice de Gini est $G = 2(0{,}5 - A) = 1 - 2A$. L'aire A est la somme des aires d'un triangle et de trapèzes de même hauteur $\frac{1}{n}$:

$$A = \frac{y_1}{2n} + \sum_{k=1}^{n-1}\frac{1}{2n}(y_{k+1}+y_k) \text{ et } y_{k+1}+y_k = \frac{1}{s}\left(2\sum_{i=1}^{k} x_i + x_{k+1}\right)$$

$$\text{d'où } A = \frac{x_1}{2ns} + \frac{1}{ns}\sum_{k=1}^{n-1}\sum_{i=1}^{k} x_i + \frac{1}{2ns}\sum_{k=1}^{n-1} x_{k+1} = \frac{1}{ns}\sum_{k=1}^{n-1}\sum_{i=1}^{k} x_i + \frac{1}{2n},$$

or $\sum_{k=1}^{n-1}\sum_{i=1}^{k} x_i = ns - \sum_{i=1}^{n} i x_i$, ainsi $A = 1 + \frac{1}{2n} - \frac{1}{ns}\sum_{i=1}^{n} i x_i$ et $G = 1 - 2A = \frac{2}{ns}\sum_{i=1}^{n} i x_i - \frac{n+1}{n}$.

En réduisant au même dénominateur, on obtient $G = \frac{1}{ns}\sum_{i=1}^{n}(2i-n-1)x_i$. Sous cette forme on voit que si toutes les retraites sont multipliées par une constante, G est invariant. On en déduit que dans un pays où les retraites sont égales à un pourcentage fixé du dernier salaire, l'indice de Gini des salaires des salariés qui prennent leur retraite est égal à l'indice de Gini des retraites de ces salariés, autrement dit le passage à la retraite ne réduit pas les inégalités. Si $n > 1$, en regroupant deux par deux les termes équidistants des extrêmes, (x_1, x_n), (x_2, x_{n-1}), ... on obtient $G = \frac{1}{ns}\sum_{i=1}^{q}(n+1-2i)(x_{n+1-i}-x_i)$ où q est la partie entière de $\frac{n}{2}$. Sous cette forme, il apparaît que G mesure bien des écarts entre les retraites et que G n'est pas invariant si toutes les retraites sont augmentées de la même valeur car bien que les différences ne changent pas, la somme s augmente, donc G diminue. Cette propriété n'est pas un inconvénient car on peut considérer que la répartition (10 000, 20 000, 30 000) est plus inégalitaire que (110 000, 120 000, 130 000) si on observe les différences relatives.



Interprétation

Étudions le cas particulier suivant : dans une population de 100 retraités, 80 ont une retraite de r euros ($r > 0$) et 20 ont une retraite de 0 euros. Le domaine situé sous la courbe de Lorenz est alors un triangle d'aire $\frac{1}{2}(0{,}8) = 0{,}4$. L'indice de Gini correspondant est $G = 1 - 2(0{,}4) = 0{,}2$. Plus généralement, dans une population où x % des retraités ont une retraite égale à zéro et les autres une retraite égale à $r > 0$, l'indice de Gini sera égal à $\frac{x}{100}$. Ainsi un indice de Gini égal à 0,3 (qui est approximativement l'indice de Gini des salaires en France) signifie que l'inégalité du système est celle d'une population où 30 % des retraités ont une retraite égale à zéro et les autres une retraite égale à $r > 0$. Ce cas particulier étant un peu trop éloigné de la réalité, on peut étudier le suivant : une population de retraités est la réunion de deux sous-ensembles disjoints E_1 et E_2 d'effectifs n_1 et n_2 ; les retraités de E_1 perçoivent une retraite égale à r_1, ceux de E_2 ont une retraite égale à r_2 ; le montant total des retraites est alors $s = n_1r_1 + n_2r_2$. On suppose que $0 < r_1 < r_2$ et $\frac{n_1}{n_1+n_2} = x > 0{,}5$ et $\frac{n_1r_1}{s} = 1 - x$: les retraités de E_1 représentent la proportion x des retraités et perçoivent la proportion $(1-x)$ du total des retraites. On montre alors que l'indice de Gini de l'ensemble des retraités est $G = 2x - 1$ soit $x = \frac{1+G}{2}$. Alors $G = 0{,}3$ signifie que l'inégalité du système est celle du système précédent dans lequel $\frac{1{,}3}{2} = 65$ % des retraités perçoivent 35 % du total des retraites.

Une autre interprétation en terme d'écarts

À retraite moyenne constante pour la population, l'indice de Gini permet aussi de mesurer l'écart moyen entre les retraites.

Cela résulte du calcul suivant : l'écart moyen entre les retraites est $e = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|$.

Si on range les retraites x_i dans l'ordre croissant, on a :

$$e = \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i) = \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j - \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i = \frac{2}{n^2}\sum_{i=1}^{n-1}\left(s - \sum_{j=1}^{i}x_j - (n-i)x_i\right)$$

ce qui donne $e = \dfrac{2(n-1)s}{n^2} - \dfrac{2}{n^2}\displaystyle\sum_{i=1}^{n-1}\sum_{j=1}^{i}x_j - \dfrac{2}{n}(s - x_n) + \dfrac{2}{n^2}\left(\displaystyle\sum_{i=1}^{n} ix_i - nx_n\right)$

or $\displaystyle\sum_{i=1}^{n-1}\sum_{j=1}^{i}x_i = ns - \sum_{i=1}^{n} ix_i$ d'où $e = \dfrac{4}{n^2}\displaystyle\sum_{i=1}^{n} ix_i - \dfrac{2(n+1)s}{n^2} = 2\left(\dfrac{s}{n}\right)G = 2mG$

ce qui permet de donner une autre interprétation de G : la moyenne des écarts entre les retraites est le produit de l'indice de Gini et du double de la moyenne des retraites. Ainsi à moyenne des retraites constante, plus l'indice de Gini est grand plus l'écart moyen entre les retraites l'est : l'indice de Gini peut détecter une augmentation de l'écart entre les retraites. Par exemple, $G = 0{,}5$ correspond à $e = m$ c'est-à-dire que la moyenne des différences des retraites est égale à la moyenne arithmétique des retraites. On peut aussi interpréter ce calcul de façon probabiliste : on choisit au hasard selon une loi uniforme avec remise deux retraités parmi les n retraités. Soit X la variable aléatoire représentant la différence de leurs retraites (en valeur absolue). L'espérance de X est la moyenne de toutes les différences possibles : $E(X) = e$.

Décomposition de l'indice de Gini

Soit $\{E_1, \ldots, E_k\}$ une partition de l'ensemble E des retraités, G l'indice de Gini de l'ensemble des retraites. Peut-on obtenir une décomposition de G en fonction des indices de Gini correspondant aux sous-ensembles $E_1, \ldots, E_k$ et d'indices mesurant l'inégalité des retraites entre deux sous-ensembles ? Pour cela on va utiliser la variable aléatoire X définie dans le paragraphe précédent et l'expression de son espérance $E(X)$ en fonction d'espérances conditionnelles. Soit X_1 et X_2 les variables aléatoires représentant les montants des retraites du premier et du second tirage. On considère la variable aléatoire $X = |X_1 - X_2|$.

Soit A_{ij} l'événement : « le premier retraité choisi appartient à E_i et le second retraité appartient à E_j ». Les A_{ij} forment une partition de $E \times E$.

On a alors $E(X) = \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{k} E\left(X|A_{ij}\right)P(A_{ij}) = \sum_{i=1}^{k} E\left(X|A_{ii}\right)P(A_{ii}) + \sum_{i\neq j} E\left(X|A_{ij}\right)P(A_{ij})$.

Si $i \neq j$, $P(A_{ij}) = \dfrac{n_i}{n} \times \dfrac{n_j}{n}$ car X_1 et X_2 sont indépendantes. Le terme $E\left(X|A_{ii}\right)$ est l'espérance conditionnelle de X sachant que les deux tirages ont été effectués dans E_i. D'après le résultat du paragraphe précédent, $E\left(X|A_{ii}\right) = 2m_iG_i$ où G_i est l'indice de Gini correspondant aux retraités de E_i et m_i la moyenne arithmétique des retraites dans e_i. Si $i \neq j$, $E\left(X|A_{ij}\right)$ est l'espérance conditionnelle de X sachant que le premier tirage a été effectué dans E_i et le second dans E_j ; soit n_i le nombre de retraités de E_i et R_i la suite des retraites des retraités de E_i ; $E\left(X|A_{ij}\right) = \dfrac{1}{n_in_j}\displaystyle\sum_{x\in R_i}\sum_{y\in R_j}|x-y|$.

Soit G_{ij} le nombre défini par $E\left(X|A_{ij}\right) = (m_i + m_j)G_{ij}$. Ainsi G_{ij} mesure bien les différences entre les retraites de E_i et de E_j et on remarque l'analogie avec la formule $E(X) = 2mG$.

On a alors $E(X) = \displaystyle\sum_{i=1}^{k} 2m_iG_i\left(\frac{n_i}{n}\right)^2 + \sum_{i\neq j}(m_i + m_j)G_{ij}\left(\frac{n_i}{n}\right)\left(\frac{n_j}{n}\right)$.

On divise par $2m$: $G = \dfrac{1}{m}\displaystyle\sum_{i=1}^{k} m_i \left(\frac{n_i}{n}\right)^2 G_i + \frac{1}{2m}\sum_{i\neq j}(m_i+m_j)\left(\frac{n_i}{n}\right)\left(\frac{n_j}{n}\right) G_{ij}$.

Le premier terme mesure les inégalités à l'intérieur de chaque groupe et le second les inégalités entre les groupes.

Exemple : soit $R = (10, 12, 14, 16, 18)$ les retraites de cinq retraités que l'on divise en deux groupes dont les retraites sont $R_1 = (10, 12)$ et $R_2 = (14, 16, 18)$. Soit G_1 et G_2 les indices de Gini de chaque groupe et G_{12} l'indice de Gini défini précédemment qui mesure l'inégalité entre les deux groupes. On trouve :

$$G_1 = \frac{1}{22};\ G_2 = \frac{1}{18};\ \sum_{x\in R_1}\sum_{y\in R_2}|x-y| = 4+6+8+2+4+6 = 30 \text{ d'où } G_{12} = \frac{30}{2\times 3\times(11+16)} = \frac{5}{27}.$$

$$G = \frac{1}{25\times 14}\left(11\times 4\times\frac{1}{22} + 16\times 9\times\frac{1}{18}\right) + \frac{2}{2\times 25\times 14}(11+16)\times 2\times 3\times\frac{5}{27} = \frac{1}{35}+\frac{3}{35} = \frac{4}{35}.$$

On peut donc dire que l'inégalité entre les groupes explique 75 % de l'inégalité de la répartition.
En utilisant la décomposition de la variance, on trouve : $v_1 = 1$; $v_2 = \dfrac{8}{3}$; $m_1 = 11$; $m_2 = 16$.

Moyenne des variances : $\dfrac{1}{5}\left(2\times 1 + 3\times\dfrac{8}{3}\right) = 2$.

Variance des moyennes : $\dfrac{1}{5}(2\times 9 + 3\times 4) = 6$.
La variance dans la population est donc 8 et on aboutit dans ce cas particulier à la même conclusion : $\dfrac{6}{8} = 75$ % de la variance s'explique par l'inégalité entre les deux groupes. La décomposition de l'indice de Gini et la décomposition de la variance ne conduisent pas toujours exactement à la même interprétation et il serait peut-être intéressant de déterminer dans quels cas particuliers les interprétations sont identiques et pourquoi pas la probabilité de tomber sur un tel cas particulier en choisissant une suite de nombres au hasard.

Conclusion

On a défini quelques indices permettant de mesurer l'inégalité d'une répartition de retraites et plus généralement d'une répartition de revenus. Il en existe bien d'autres. L'indice de Gini et l'indice $\dfrac{V}{V_{\max}}$ sont tous deux des nombres compris entre 0 et 1 d'autant plus grands que l'inégalité est forte mais pour comparer les inégalités de deux ou plusieurs répartitions il faut bien sûr utiliser le même indice. Par exemple, il serait imprudent de dire qu'une répartition ayant un indice de Gini de 0,3 est plus inégalitaire qu'une répartition ayant un indice $\dfrac{V}{V_{\max}}$ de 0,2 car ces deux indices ne mesurent pas la même forme d'inégalité. On peut faire une remarque analogue pour la décomposition de la variance et celle de l'indice de Gini.

Pierre Carriquiry est aujourd'hui à la retraite ; il a enseigné à l'École Nationale de Commerce. Il est membre de l'APMEP depuis 35 ans.

pierre.carriquiry@free.fr

Sommaire du nº 538

Mathématiques à l'oral

CRAP-Cahiers pédagogiques

CultureMATH

MATh.en.JEANS

APMEP

www.apmep.fr